

# Mass spectrometry-based proteomic analysis of PDAC samples

**Do Young Hyeon**

Seoul National University <https://orcid.org/0000-0002-1825-0097>

**Dowoon Nam**

Korea University

**Youngmin Han**

Seoul National University College of Medicine

**Duk Ki Kim**

Seoul National University College of Medicine

**Gibeom Kim**

Seoul National University

**Daeun Kim**

Ajou University

**Jingi Bae**

Korea University

**Seunghoon Back**

Korea University

**Dong-Gi Mun**

Korea University

**Inamul Hasan Madar**

Korea University

**Hangyeore Lee**

Korea University

**Su-Jin Kim**

Korea University

**Hokeun Kim**

Korea University

**Sangyeop Hyun**

Ajou University

**Chang Rok Kim**

Seoul National University

**Seon Ah Choi**

Seoul National University

**Yong Ryoul Kim**

Seoul National University

**Juhee Jeong**

Seoul National University College of Medicine

**Suwan Jeon**

Seoul National University College of Medicine

**Yeon Woong Choo**

Seoul National University College of Medicine

**Kyung Bun Lee**

Seoul National University College of Medicine

**Wooil Kwon**

Seoul National University College of Medicine

**Seunghyuk Choi**

Hanyang University

**Taewan Goo**

Seoul National University

**Taesung Park**

Seoul National University

**Young-Ah Suh**

Seoul National University College of Medicine

**Hongbeom Kim**

Seoul National University College of Medicine

**Ja-Lok Ku**

Seoul National University College of Medicine

**Min-Sik Kim**

Daegu Gyeongbuk Institute of Science and Technology

**Eunok Paek**

Hanyang University

**Daechan Park** (✉ [dpark@ajou.ac.kr](mailto:dpark@ajou.ac.kr))

Ajou University

**Keehoon Jung** (✉ [keehoon.jung@snu.ac.kr](mailto:keehoon.jung@snu.ac.kr))

Seoul National University College of Medicine

**Sung Hee Baek** (✉ [sbaek@snu.ac.kr](mailto:sbaek@snu.ac.kr))

Seoul National University

**Jin-Young Jang** (✉ [jangjy4@snu.ac.kr](mailto:jangjy4@snu.ac.kr))

Seoul National University College of Medicine

**Daehee Hwang** (✉ [daehee@snu.ac.kr](mailto:daehee@snu.ac.kr))

Seoul National University

**Sang-Won Lee** (✉ [sw\\_lee@korea.ac.kr](mailto:sw_lee@korea.ac.kr))

Korea University

---

## Method Article

**Keywords:**

**Posted Date:** January 19th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.pex-2063/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The associated publication reports proteogenomic analysis of human pancreatic ductal adenocarcinoma (PDAC), where we provided significantly mutated genes (SMGs)/biomarkers, cellular pathways, and cell types as potential therapeutic targets to improve stratification of patients with PDAC. This protocol describes the detailed methods for mass spectrometry-based proteomic analysis of PDAC samples, including sample preparation, liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, and data analyses (peptide/protein identification and quantitation).

## Introduction

## Reagents

## Equipment

## Procedure

### Tissue homogenisation

To prevent blood contamination, the cancer tissues were briefly (ca. <3 min) washed with phosphate-buffered saline on ice. The washed tissues were then individually placed in a cryovial (430487, Covaris) before they were frozen in liquid nitrogen. Once on the dry ice, each of the frozen tissues was transferred to a TT05 tissue bag (520140, Covaris) and placed into liquid nitrogen again for 30 s before being pulverised at level 2 using a Cryoprep device (CP02, Covaris). We used a TT1 tissue bag (520001, Covaris) for pulverisation at level 3 when the tissue weight was more than 50 mg.

### Protein extraction and digestion

A portion of each tissue powder sample was transferred to a new EP tube and mixed with lysis buffer (4% SDS, 0.1 M Tris-HCl pH 7.6, and phosphatase inhibitor [04906837001, Roche]) and then lysed using a probe sonicator (Q55 Sonicator, Qsonica) for 30 s on ice. The lysis process was repeated five times. The lysate was centrifuged at 16,000 *g* at 20 °C for 10 min, and the supernatant was transferred to a new EP tube. The protein concentration in the lysate was measured using a BCA assay (23225, Thermo Scientific).

Each of the 500 µg protein from the individual tissues was digested by the filter-aided sample preparation method with a slightly modified protocol that has been described previously<sup>1,2</sup>. Briefly, the proteins were reduced with SDT buffer (4% SDS in 0.1 M Tris-HCl, pH 7.6, and 0.1 M DTT) at 37 °C for 45 min with shaking at 300 rpm and then boiled for 10 min at 95 °C on a thermomixer (Comfort, Eppendorf). The reduced proteins were sonicated for 10 min using a bath sonicator (Branson 5800, Branson

Ultrasonics) and centrifuged at 16,000 *g* for 5 min before being transferred to a Microcon YM-30 filter (MRCF0R030, Millipore Corporation). The sample was centrifuged at 14,000 *g* for 30 min and mixed with 200  $\mu\text{L}$  of 8 M urea in 0.1 M Tris-HCl at pH 8.5. The sample was then centrifuged at 14,000 *g* for 60 min at 20 °C to remove the detergent. After repeating these steps thrice, the proteins were alkylated with 100  $\mu\text{L}$  of 50 mM iodoacetamide in 8 M urea at room temperature in the dark for 25 min and centrifuged at 14,000 *g* for 30 min. The filter was washed four times with 200  $\mu\text{L}$  of 8 M urea and then twice with 100  $\mu\text{L}$  of 50 mM  $\text{NH}_4\text{HCO}_3$ . After the buffer exchange, trypsin (V5111, Promega) was added at an enzyme-to-protein ratio of 1:50 (w/w) and then placed in a thermomixer for overnight incubation at 37 °C. The second digestion was carried out for 6 h with additional trypsin (enzyme-to-protein ratio of 1:100). The peptide samples were then eluted by centrifugation at 14,000 *g* for 20 min, and 60  $\mu\text{L}$  of 50 mM  $\text{NH}_4\text{HCO}_3$  was added and rinsed twice at 14,000 *g* for 30 min. The peptide concentrations were measured using a BCA assay (23225, Thermo Scientific) before vacuum-drying (concentrator plus, Eppendorf) and stored at 80 °C.

### **Tandem mass tag labelling and desalting**

Peptides (200  $\mu\text{g}$ ) from each of the nine different tumour tissues and one universal reference were labelled using the 10-plex tandem mass tag (TMT) reagent. The universal reference, which was prepared by pooling 50  $\mu\text{g}$  of peptides obtained from 189 PDAC tissues, was used as a representative peptide sample of PDAC. Each peptide sample (100  $\mu\text{g}$ ) was resuspended in 100  $\mu\text{L}$  of 100 mM triethylammonium bicarbonate (TEAB) using sonication and vortexing. Anhydrous acetonitrile (40  $\mu\text{L}$ ) was added to the TMT reagent and transferred to the sample tubes. The tubes were incubated at room temperature for 1 h and then quenched with 8  $\mu\text{L}$  of 0.5% hydroxylamine. After 15 min of incubation at room temperature for the quenching reaction, each of the 10 TMT channels was pooled and vacuum-dried.

The labelled peptides (2 mg) were dissolved in 900  $\mu\text{L}$  of 10 mM TEAB and desalted. An Agilent 1260 Infinity HPLC system (Agilent, Santa Clara, CA) with an Xbridge C18 guard column (4.6  $\times$  20 mm, 130 Å, 5  $\mu\text{m}$ ) and an analytical column (4.6  $\times$  250 mm, 130 Å, 5  $\mu\text{m}$ ) were used for desalting. After 900  $\mu\text{L}$  of the sample was injected, the gradient of solvent A (10 mM TEAB in water, pH 7.5) and solvent B (10 mM TEAB in 90% acetonitrile) was used as follows: 100% solvent A for 25 min, 0–70% solvent B for 5 min, and then 70% solvent B for 30 min. Eluents were collected every minute from start to end, and peptide samples were pooled from the collected well plate according to the UV spectrum. Pooled peptides were dried in a vacuum concentrator. For the global proteome analysis, TMT peptides (100  $\mu\text{g}$ ) were used. The remaining TMT peptides were subjected to immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment. Using the same sample amount for all injections, peptide abundance was normalized relative to that of total peptide amount in individual samples, preventing the selection of many protein signatures for subtypes that produced high protein amounts. This normalization ensures

the robustness of subsequent analyses (e.g. clustering and signature identification) with respect to variations in the translation rate, with little loss of sensitivity in the analyses.

### **One-pot phosphopeptide enrichment and global peptide preparation**

We performed one-pot phosphopeptide enrichment, wherein IMAC phosphopeptide enrichment was performed for the entire TMT-labelled peptides before fractionation. To enrich the phosphopeptides from the entire TMT-labelled peptides, we used a previously described IMAC method<sup>3</sup>. Briefly, Ni-NTA magnetic beads (1 mL; 36111, Qiagen) were washed three times with 800  $\mu$ L of nano-pure water and then treated with 800  $\mu$ L of 100 mM EDTA for 30 min on end-over-end rotation (SB3, Stuart). After removing the EDTA solution, the magnetic beads were again washed thrice with 800  $\mu$ L of nano-pure water and reacted with 800  $\mu$ L of 10 mM FeCl<sub>3</sub> for 30 min on end-over-end rotation. Fe<sup>3+</sup>-NTA beads were washed three times with 800  $\mu$ L of nano-pure water and resuspended in 800  $\mu$ L of 1:1:1 acetonitrile/methanol/0.01% acetic acid for aliquoting the bead solution into two 5 mL EP tube (0030119401, Eppendorf), each of which contained 400  $\mu$ L of the bead solution.

The TMT-labelled peptides (2 mg) were resuspended in 4 mL of washing buffer (20% water, 0.1% TFA in acetonitrile), and the aliquoted beads were washed with 1.6 mL of the same washing buffer. After removing the washing buffer from the beads, the resuspended peptides were added and incubated for 30 min with an end-over-end rotator. The supernatant was collected, and the beads were washed with 2 mL of washing buffer four times. Each supernatant was collected for analysis as a global proteome. Phosphopeptide elution was performed using 1.5 mL of 1:1 acetonitrile/2.5% ammonia in 2 mM phosphate buffer after incubating for 1.5 min. The suspension was transferred to a new tube and acidified with 10% TFA to reach pH 3.5–4.0 before subsequent desalting.

The phosphopeptides were vacuum-dried and resuspended in 75  $\mu$ L of 0.5% TFA in water to perform spin column desalting. The spin column (744614, Harvard apparatus) was first washed with 75  $\mu$ L of 0.1% TFA in acetonitrile/water (80:20 v/v) by centrifuging at 200 *g* for 1 min. The column was equilibrated with 0.1% TFA in water two times at 200 *g* and two times at 400 *g* for 1 min each. The resuspended sample was added and centrifuged at 200 *g* for 1–2 min for binding. Flow-through was again added to the column for rebinding at 200 *g* for 1–2 min. The salt was removed by adding 75  $\mu$ L of 0.1% TFA in water and centrifuged at 200 *g* for 1 min. This step was repeated thrice. The desalted sample was then eluted with 75  $\mu$ L of 80% acetonitrile and 0.1% TFA washing buffer and then centrifuged at 100 *g* for 1 min. This step was repeated one more time before 75  $\mu$ L of 100% acetonitrile was added and centrifugation was performed at 200 *g* for 1 min. Finally, the washing buffer (75  $\mu$ L) was added and the sample was centrifuged at 200 *g* for 10 s and at 1,000 *g* for 50 s to ensure high recovery efficiency. Each eluate was pooled and vacuum-dried. Both phosphopeptides and global peptides were stored at 80 °C.

## Online 2D non-contiguous fractionating and concatenating-RP/RPLC experiments

Both the global peptide and phosphopeptide samples were fractionated using the previously developed online non-contiguous fractionating and concatenating (NCFC) fractionator<sup>4,5</sup> coupled with a dual online RPLC system. The NCFC fractionator was modified to generate 24 fractions. Briefly, the NCFC fractionator consists of a 6-port valve (6-port/3-channel/2-position, C72MX-4676, VICI) equipped with a mid-pH RP column (150  $\mu\text{m}$   $\times$  50 cm) and two 25-port valves (25-port/1-channel/multi-position, C5M-66024D, VICI) that were interconnected by 24 NCFC fraction loops (200  $\mu\text{m}$   $\times$  22.3 cm) to form an NCFC fractionator. Two sets of SPE columns (150  $\mu\text{m}$   $\times$  3 cm) and analytical columns (75  $\mu\text{m}$   $\times$  100 cm) were installed on two 6-port valves (6-port/3-channel/2-position, C72MX-4676, VICI). A mid-pH RP column, two SPE columns, and two analytical columns were manufactured in-house by slurry packing of Jupiter C18 particles (Jupiter, 3  $\mu\text{m}$ , 300  $\text{\AA}$ , Phenomenex)<sup>6</sup>, and both mid-pH columns and two analytical columns were maintained at 60  $^{\circ}\text{C}$  using a column heater (MCT200, Analytical Sales & Product)<sup>7</sup>. The DO-NCFC-RP/RPLC system was operated using three binary pumps (Ultimate 3000 NCP-3200RS, Dionex, Thermo Fisher Scientific) and an autosampler (Ultimate 3000 WPS-3000TPL RS, Dionex, Thermo Fisher Scientific).

For the 1<sup>st</sup> dimensional separation, solvent A (10 mM TEAB in water, pH 7.5) and mid solvent B (10 mM TEAB in 99 % ACN, pH 7.5) were used as the mobile phases. For the global proteome analysis, 100  $\mu\text{g}$  of peptide was injected and separated using a 120 min gradient (1–50% mid solvent B over 120 min with 1  $\mu\text{L}/\text{min}$  flow rate). Two 25-port selector valves were synchronously step-rotated every minute so that eluates from the 1<sup>st</sup> dimension column were subsequently deposited in each of the 24 fraction loops in turn. After the first collection cycle, the NCFC fractionator continued to step-rotate such that the 25<sup>th</sup> eluate was concatenated with the 1<sup>st</sup> eluate, the 26<sup>th</sup> with the 2<sup>nd</sup>, and so on. After five full collection cycles, each of the 24 fraction loops contained five non-contiguous eluates from the 1<sup>st</sup> dimensional separation. For the phosphoproteome analysis, the same method was used except that two-step rotations (e.g. stepping position 1  $\rightarrow$  3  $\rightarrow$  5 ...  $\rightarrow$  24  $\rightarrow$  1) were used to produce 12 NCFC fractions with a flow rate of 0.5  $\mu\text{L}/\text{min}$ .

The NCFC fractionator was intercalated within a dual-online reverse-phase liquid chromatography system (DO-RPLC)<sup>4</sup> to form a DO-NCFC-RP/RPLC system. The DO-RPLC system combines two sets of analytical columns and two SPE columns, which operate in parallel and alternately, to achieve a 100% duty cycle (separation on one column and equilibration on the other column, simultaneously). After the online NCFC fractionation was completed, each fraction in the 24 NCFC loops was acidified and diluted 10 times with a buffer (0.2% TFA in water) while transferring to the SPE column for the 2<sup>nd</sup> dimensional separation. To perform the 2<sup>nd</sup> dimensional reverse-phase separation, solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile) were used through the analytical column with the following gradient: 3% solvent B with 1  $\mu\text{L}/\text{min}$  flow rate for 0–0.2 min, 3.2–4% with 0.05–0.3  $\mu\text{L}/\text{min}$  for 0.2–1.2 min, 4–7% with 0.3  $\mu\text{L}/\text{min}$  for 1.2–5 min, 7–35% with 0.25  $\mu\text{L}/\text{min}$  for 5–160 min, 35–80% with

0.25–0.4  $\mu\text{L}/\text{min}$  for 160–165 min, 80% with 0.4  $\mu\text{L}/\text{min}$  for 165–175 min, and 3% with 0.4  $\mu\text{L}/\text{min}$  for 175–180 min.

## Mass spectrometry analysis

A quadrupole-orbitrap mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) at the Center for Proteogenome Research was used to analyse both the global proteome and phosphoproteome. Eluted peptides from the LC system were ionized by electrospray ionisation with an electric potential of 2.4 kV and desolvation capillary temperature of 250 °C. The MS1 scan was performed in the range of 400–2,000 Th for 60,000 resolution with a maximum ion injection time of 20 ms. The ten most abundant ions were selected for the MS2 scan with a 1.2 Th isolation window. The MS2 resolution was 45,000 with an 86 ms maximum ion injection time and normalised collision energy of 32 for higher-energy collision dissociation. The first fixed  $m/z$  was 120 Th, dynamic exclusion time was 25 s, charge state of 1 was discarded, and both MS1 and MS2 automatic gain control target values were  $5 \times 10^6$ .

## Generation of sample-specific customised database

A sample-specific customised database was constructed with a slight modification to the previously reported method<sup>8</sup>. Briefly, the customised database included two databases of expressed transcripts and variants of each patient. To generate the expressed transcript database, we first used STAR (version 2.5.3)<sup>9</sup> to map the reads obtained from the mRNA sequencing to the reference genome, human GRCh38, with the default parameters, and then used RSEM (version 1.3.0)<sup>10</sup> to estimate the transcript abundance in fragments per kilobase of transcript per million mapped reads (FPKM). For each patient, the expressed transcripts were defined as those of the non-zero FPKM. We then translated the expressed transcripts into amino acid sequences as previously described<sup>8</sup>. To generate the variant database, we obtained the genomic variants identified from exome sequencing data; the single amino acid variations were translated to protein sequences based on the Ensembl transcript model as previously described<sup>8</sup>. Briefly, peptide sequence translation starts at each SNV and INDEL position and is extended into both directions, allowing up to three mis-cleavages, while minimising the redundant sequence generation by merging overlapping peptides. If the length was less than 50 amino acids, the translation was extended to the first tryptic terminus after the 50<sup>th</sup> amino acid position. If a mutation call fell upon a STOP codon, the translation extended the coding sequence (CDS) of the transcript up to 20 amino acids until another STOP codon was present or mRNA sequencing evidence for C-terminal extension diminished below a read count of three. To minimise the redundancy in the resulting amino acid sequences, translation from the CDS of a transcript to an amino acid sequence was applied only to the local CDS region near the variant positions.

A multiplexed protein database was constructed by combining the customised database of nine patients that were multiplexed in a TMT experiment with the removal of redundant entries. The multiplexed database was subsequently combined with the SwissProt (version 2018. 01, 42,271 entries) database with redundancy removal to construct the final multiplexed unified protein database.

## Multi-stage database search

MS/MS data were analysed using the multi-stage database search method as previously reported<sup>8</sup>. Briefly, MS/MS data were first processed with mPE-MMR<sup>11</sup> to correct and refine the precursor ion mass, especially by assigning multiple precursor masses of the multiplexed spectra. The resultant MS/MS data were first subjected to MS-GF+ (version 9949)<sup>12</sup> database search using the corresponding multiplexed unified protein database. The unidentified MS/MS data from the 1<sup>st</sup> stage search were subsequently analysed using MODplus for blind modification search<sup>13</sup>.

For the 1<sup>st</sup> stage search, a semi-tryptic option and a precursor mass tolerance of 10 ppm were used. In the database search of global proteome data, TMT on N-terminus and lysine and carbamidomethylation on cysteine were set as static modifications, whereas methionine oxidation and asparagine and glutamine deamidation were set as variable modifications. Similar modifications with the addition of serine, threonine, and tyrosine phosphorylation were applied to the phosphoproteome data analysis. After the database searches were performed, the resultant peptide-spectrum matches (PSMs) from a fraction were mapped through unique mass classes (UMCs) of the corresponding fraction by their scan numbers; the PSM with the highest search score in a UMC was chosen as the representative PSM for that UMC. All representative PSMs from all 24 fractions were subsequently bifurcated to each patient by checking the corresponding protein IDs of the peptides in each customised protein database. For each patient, all peptides with the highest search scores were subsequently selected for target-decoy analysis, and the target peptides filtered by an FDR of 1% at the peptide level (PepFDR) were finally obtained. Dual target-decoy distributions for the SwissProt peptides and customised database peptides were employed for each patient's data, as previously described<sup>8</sup>. Similar phosphopeptide bifurcations were obtained as described above. For the 2<sup>nd</sup> stage MODplus search, 59 variable modifications and three static modifications (TMT on N-terminus and lysine and carbamidomethylation on cysteine) were used. MODplus search results were filtered using processes similar to those used for MS-GF+ search results with 1% PepFDR.

For the phosphopeptide data, the unidentified MS/MS data after the 2<sup>nd</sup> stage search were further subjected to spectral library search<sup>3</sup>. We constructed a spectral library for phosphopeptides identified from the phosphoproteome of 150 PDAC tissues. The spectral library contained the experimental mass, normalised elution time, and representative MS/MS spectra of the observed phosphopeptides. Among the PSMs acquired from the 1<sup>st</sup> and 2<sup>nd</sup> stage searches, the PSMs with the highest search scores from each charge state of the phosphopeptides were selected as the representative spectra of the individual

peptides. The experimental retention time was converted to the normalised elution time using the iRT peptide (Ki-3002-2, Biognosys) spiked with each NCFD fraction. We identified a total of 20,7762 non-redundant phosphopeptides across 150 PDAC tissues and compiled a total of 281,570 representative spectra in the spectral library. To construct the decoy spectral library database, each peptide sequence from the spectral library database was shuffled to form decoys with the C-terminal arginine and lysine fixed<sup>14</sup>. Moreover, a spectrum-spectrum match FDR of 1% was used, and an additional cut-off value of the Excrr of six or higher was used<sup>3</sup>.

## **Protein grouping and genome mapping**

The identified non-redundant peptides were used to obtain protein groups by bipartite graph analysis using an in-house program, as described previously<sup>3</sup>. Briefly, among the component proteins of each protein group, the protein with the highest number of associated peptides was selected as the representative of that protein group. When multiple proteins had the same number of non-redundant peptides, the protein with the higher protein sequence coverage was chosen as the representative protein. Additionally, a protein with unique peptides was selected as a separate representative protein. For the global proteome analysis, two or more sibling non-redundant peptides were required for the protein group identification. For phosphopeptide analysis, all protein groups identified by the phosphorylated peptides were used. Then, we mapped all representative proteins from a tissue sample to the UniProt mapping table (version 2018.01) to obtain non-redundant gene IDs. To obtain protein-coding genes identified by mRNA sequencing data, transcripts with FPKM  $\geq 1$  were used.

## **Peptide and protein quantitation**

We used 10-plex TMT labelling to quantify the protein abundance of nine tumour tissues (channels 126-130C) along with the universal reference (channel 131). After isotope impurity correction, the intensities of 10 TMT reporter ions (126-131) were extracted from all MS/MS scans with a mass tolerance of 0.005 Da, and then normalised using quantile normalisation<sup>15</sup>. For quantification, we used the precursor isolation purity (PIP) value, which is the ratio of the intensity of the precursor peaks to the total peak intensity within the isolation window. Among the MS/MS spectra corresponding to a peptide (or phosphopeptide), the MS/MS spectrum with a PIP value  $> 70$  was used for quantifying the peptide. When multiple PSMs had a PIP value  $> 70$ , the median intensity of the PSMs was used as the intensity of the peptide. The  $\log_2$  fold-change of a peptide between each of the nine tumour tissues and the universal reference was calculated as the  $\log_2$  ratio of each of the tumour reporter ion intensities and the universal reporter ion intensity. In the global proteome, the  $\log_2$  fold-change of the proteins was estimated as the median  $\log_2$  fold-change of all non-redundant peptides corresponding to a protein. Proteins with at least

two non-redundant peptides were selected for the subsequent analyses (e.g. mutation-phosphorylation, mRNA-protein correlation analyses, and clustering).

Before proceeding with the analyses, we corrected for batch effects in  $\log_2$  fold-change of proteins and phosphopeptides between different TMT sets. For the 17 TMT sets, the quantification procedure described above generated an  $n \times 150$  matrix containing  $\log_2$  fold-changes of phosphopeptides (or proteins) for the 150 samples analysed.  $n$  is the maximum number of phosphopeptides (or proteins) measured in the 17 TMT sets. To correct for batch effects, we first concatenated the  $\log_2$  fold-changes for all phosphopeptides (or proteins) in the samples from each TMT set into a single column. After merging the concatenated columns for all TMT sets into a 17-column matrix, we performed quantile normalisation<sup>15</sup> so that the distributions of  $\log_2$  fold-changes across all 17 TMT sets would be the same. The normalised  $\log_2$  fold-changes were then returned to their original positions in the  $n \times 150$  matrix. Then, distributions across different TMT sets were matched to retain the ranks of  $\log_2$  fold-changes for phosphopeptides (or proteins) in each TMT set while making  $\log_2$  fold-changes with the same ranks (or quantiles) equal across all TMT sets. As a result, all statistical moments (e.g. mean, variance, and skewness) became the same across all TMT sets. This normalisation approach provides statistical stability in subsequent analyses by effectively removing systematic biases from the data derived from different batches<sup>15</sup>. After correcting for the major batch effect across TMT sets, we finally performed quantile normalisation on the normalised  $n \times 150$  matrix to match the distributions of  $\log_2$  fold-changes of phosphopeptides (or proteins) in each sample to further eliminate any systematic bias for individual samples.

## Troubleshooting

## Time Taken

## Anticipated Results

## References

- 1 Jo, D. H. *et al.* Quantitative Proteomics Reveals beta2 Integrin-mediated Cytoskeletal Rearrangement in Vascular Endothelial Growth Factor (VEGF)-induced Retinal Vascular Hyperpermeability. *Mol Cell Proteomics* **15**, 1681-1691, doi:10.1074/mcp.M115.053249 (2016).
- 2 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).
- 3 Mun, D. G. *et al.* Proteogenomic Characterization of Human Early-Onset Gastric Cancer. *Cancer Cell* **35**, 111-124 e110, doi:10.1016/j.ccell.2018.12.003 (2019).

- 4 Lee, H. *et al.* A simple dual online ultra-high pressure liquid chromatography system (sDO-UHPLC) for high throughput proteome analysis. *Analyst* **140**, 5700-5706, doi:10.1039/c5an00639b (2015).
- 5 Lee, H. *et al.* Efficient Exploitation of Separation Space in Two-Dimensional Liquid Chromatography System for Comprehensive and Efficient Proteomic Analyses. *Anal Chem* **88**, 11734-11741, doi:10.1021/acs.analchem.6b03366 (2016).
- 6 Lee, J. H. *et al.* Fully automated multifunctional ultrahigh pressure liquid chromatography system for advanced proteome analyses. *J Proteome Res* **11**, 4373-4381, doi:10.1021/pr3004166 (2012).
- 7 Hyung, S. W., Kim, M. S., Mun, D. G., Lee, H. & Lee, S. W. The effect and potential of using a temperature controlled separation column with ultra-high pressure microcapillary liquid chromatography/tandem mass spectrometry on proteomic analysis. *Analyst* **136**, 2100-2105, doi:10.1039/c0an00724b (2011).
- 8 Park, H. *et al.* Compact variant-rich customized sequence database and a fast and sensitive database search for efficient proteogenomic analyses. *Proteomics* **14**, 2742-2749, doi:10.1002/pmic.201400225 (2014).
- 9 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 10 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
- 11 Madar, I. H. *et al.* Multiplexed Post-Experimental Monoisotopic Mass Refinement (mPE-MMR) to Increase Sensitivity and Accuracy in Peptide Identifications from Tandem Mass Spectra of Cofragmentation. *Anal Chem* **89**, 1244-1253, doi:10.1021/acs.analchem.6b03874 (2017).
- 12 Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**, 5277, doi:10.1038/ncomms6277 (2014).
- 13 Na, S., Kim, J. & Paek, E. MODplus: Robust and Unrestrictive Identification of Post-Translational Modifications Using Mass Spectrometry. *Anal Chem* **91**, 11324-11333, doi:10.1021/acs.analchem.9b02445 (2019).
- 14 Lam, H., Deutsch, E. W. & Aebersold, R. Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics. *J Proteome Res* **9**, 605-610, doi:10.1021/pr900947u (2010).
- 15 Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193, doi:10.1093/bioinformatics/19.2.185 (2003).